

Sequence Listing

Individual Applicant

Street :
 City : Athens
 State : GA
 Country : USA
 PostalCode : 30602-7272
 PhoneNumber : 706-542-0928
 FaxNumber : ____-____-____
 EmailAddress : wparrott@uga.edu
 <110> LastName : Parrott
 <110> FirstName : Wayne
 <110> MiddleInitial : A
 <110> Suffix : PhD.

Individual Applicant

Street :
 City : Athens
 State : GA
 Country : USA
 PostalCode : 30602-7272
 PhoneNumber : 706-542-0928
 FaxNumber : ____-____-____
 EmailAddress :
 <110> LastName : LaFayette
 <110> FirstName : Peter
 <110> MiddleInitial : R
 <110> Suffix : PhD.

Individual Applicant

Street : 1907 South Milledge Ave
 City : Athens
 State : GA
 Country : USA
 PostalCode : 30605-____
 PhoneNumber : ____-____-____
 FaxNumber : ____-____-____
 EmailAddress :
 <110> LastName : Kane
 <110> FirstName : Patrick
 <110> MiddleInitial : M
 <110> Suffix :

Application Project

<120> Title : Arabitol or Ribitol As Positive Selectable Markers
 <130> AppFileReference : UGA-855R
 <140> CurrentAppNumber :
 <141> CurrentFilingDate : ____-__-__

Sequence

<213> OrganismName : Escherichia coli
 <400> PreSequenceString :
 atgaacgaac aatttacatg gctgcacatc gggtaggtt cttttcatcg cgcacatcag 60
 gcgtgggtatc tacaccgttt gcaggtgatg ggcgataaac gctggagcat tgctgcgggc 120
 aatattcgta atgatgctga acatgtcgtc caggcactca gtgcacagaa aggtcgctat 180
 gtgctggaaa ccgtcagccc ggaaggggta agcgaatatg aagagatcac ctcaattcag 240

aagttgatac	cgtggcaggc	agatttacaa	ccgctgattg	ctgaaggggc	agatccgaag	300
acaaaagtga	ttgctttcac	cgtcaccgaa	ggcgggtact	acctgaatac	cagtcacaaa	360
ctggaagtta	acaatcctga	tttagcggca	gatcttaaag	ggggatgcaa	aacaatttac	420
ggtgttatta	cccgtatcct	cgaagcgcgt	atggcaaata	acgccggacc	actaaccctg	480
atgaacgaac	aatttacatg	gctgcacatc	gggttaggtt	cttttcatcg	cgcacatcag	540
gcgtggatc	tacaccgttt	gcaggtgatg	ggcgataaac	gctggagcat	tgctcggggc	600
aatatctcga	atagatcgta	acatgctgta	caggcactca	gtcgacagaa	aggtcgctat	660
gtgctgga	ccgtcagccc	ggaaggggta	agcgaatatg	aagagatcac	ctcaattcag	720
aagttgatac	cgtggcaggc	agatttacaa	ccgctgattg	ctgaaggggc	agatccgaag	780
acaaaagtga	ttgctttcac	cgtcaccgaa	ggcgggtact	acctgaatac	cagtcacaaa	840
ctggaagtta	acaatcctga	tttagcggca	gatcttaaag	ggggatgcaa	aacaatttac	900
ggtgttatta	cccgtatcct	cgaagcgcgt	atggcaaata	acgccggacc	actaaccctg	960
ctgaagtgcg	ataacgtgcg	ccataatggt	gaacgtttcc	atgatggcct	ggttgagttt	1020
ctccagctaa	ctggcaca	ggatgtcatc	gactggctga	gtacaaatac	cacttgcccg	1080
aataccatgg	ttgaccgat	tacgcctcgt	ccggcagcag	aacttcgggc	acggatcaag	1140
gctcaaacgg	gtattgccga	taaagcgccg	gtaatgggcg	aaacctttat	ccagtgggtc	1200
gtggaagata	atttcctgta	tgtccgtccg	gcactggaga	aggtcgggtg	cgaactgggtg	1260
gcgtcggtaa	tcccctatga	agaggcgaag	attcgcattc	ttaactcttc	acacagttgc	1320
atcgctggg	caggtacgtt	aatcggtc	aaataatacc	acgaaagcac	aatgaccgat	1380
tttatctatc	agatttccga	ccgctacgtg	ctggaagatg	tcattctctg	cttgggcgat	1440
aacgtatc	atttgccaac	ctaccgtgat	gtgttactca	agcgttttac	caatccacat	1500
attcaggaca	ccaaccaacg	cgtcgtcgcg	gatggtttct	cgaaaattcc	ggcgtatgatt	1560
gccccacac	tgcgagagt	ctaccagcga	ggcgttcgcc	cgaatgccac	cgccatgtta	1620
cctgcactgt	tttacgtatt	catggagcag	tggcatcacg	gcaaactgcc	ctatgaatat	1680
caggatggca	tccttgatgc	accagctgtc	catgcaatgt	tacagttcgc	cgatcccgtc	1740
gcgtgtttatg	ccagtgataa	agcgtcgttt	ggcgatttaa	ccgaacgtga	agattttgcc	1800
gcgtgtttgc	gcgaataaat	ccctgacgtc	tacgcgttaa	ttaactaa		1848

<212> Type : DNA

<211> Length : 1848

SequenceName : SEQ ID NO 1

SequenceDescription : arabitol dehydrogenase

Custom Codon

Sequence Name : SEQ ID NO 1

Sequence

<213> OrganismName : Escherichia coli

```
<400> PreSequenceString :
```

atcgattgag	cagtttgcct	cacacggcag	ctaaattccc	gttcagtgcg	tgcaagcaaa	60
cagagactat	aaattcgcgc	tggtaaaagg	atttatatgat	gaatcactct	gtgccctcta	120
tgaatactcc	ccttaatggc	aaagttgcag	ctatcactgg	cgctgcgtca	ggtattggcc	180
tgcaattgtc	aaaaacgctg	ctcgcgtcag	gagcaaaagg	agtattgatt	gaccgggaag	240
gcgcaaaact	gcacaagatt	gtcgcgtagt	tggcgaaaa	cgcgtacgcg	ctgcaactcg	300
atctcttcaa	taatcagcaa	gtcgataaca	tactgcgcca	cattatcgaa	ctggcgggtg	360
ggctggatat	ttttcatgcc	aatgcaggcg	cttatattgg	cggcccagtg	gctgaagggtg	420
atccagatgt	ctgggatcgt	gtgttaaata	tgaatataaa	tgcggcgttt	cgctgtgtcc	480
gtgcagtcct	gccgcataatg	attgcgcaga	ggtcgggcga	tataattttt	accagttcca	540
tcgcgggcgt	cgttcgggtt	atctgggaac	cgatctacac	cgcgtccaaa	tttgccggtc	600
aggcattcgt	acacactacc	cgcgcgcagg	tttctcaata	tggcgtgcgt	gtgggtgcgg	660
tgctgccagg	accagtagtc	actgcctcgt	ttgatgactg	gccaaaagcc	aaaattggaag	720
aagccctggc	aaatggtagc	ctgatgcaac	cgattgaagt	ggcggaatca	gtattgttta	780
tggtgacctg	ctcgaaaaat	gtcacctgtc	gagatttagt	gatcctgcct	ggcagtgctg	840
atctgtaagg	gcgcaatcat	gacaataacc	aaaacggtta	ttggtgttga	tgtgggatca	900
ggcagtgctc	gcgcggggat	ttttgatctc	aacggatctc	tgctatccca	tgccacagaa	960
aaaatacacg	ctacgcggcg	cagcgggaag	cgcgtggaac	agtccagcca	ggagatctgg	1020
caggcggctc	gttcatgtat	tcgtaatgcg	ctcactctgg	cagacggttg	tgacaaaagt	1080
gtggcaggca	tcgggttttg	tgccacctgt	tctctggtgg	tactggataa	aaacggtgat	1140
ccattgcctg	tcagcccggg	aggagatgca	aagcaaaata	tcattgtgtg	gatggatcac	1200
cgcgccaccg	aacaagcaga	gcgaatcaat	gccactcacc	atccggtgct	gaactacgtc	1260
ggtgggtaaaa	tttcgcctga	aatggaaaaa	cggaaaaatt	cttggtcgaa	agaaaatatg	1320
ccagagatct	acgaacgtgc	cggacaattt	ttcgcattcg	cctgatttct	gacctggcgg	1380
cgtaaccggtg	attttaqcgq	ttcagtatgc	actgttacct	gtaaatggac	gtggctggca	1440

```

catgaaaatc gctgggatcc agattatttc cgcaccattg gccttgacga gttagcggat 1500
gaagatttta ttcgcattgg tcatcatatt gtttctcccg gaacaccttg tggaaatggg 1560
ttaacagcac aagccgcggc agagatggga ttactccccg gcacacctgt cgtgttaggg 1620
ttaattgatg ctcacgctgg tggcatcggt acggttggcg tagaagggtg agcgtgaac 1680
aatctcgctg atgttttcgg cacttcttca tgcaccatgg catctaccac ttctccctcg 1740
tttgtaccgg gtgtctgggg gccgtattac agtgcgatgg ttccagggtc gtggttagtt 1800
gaaggcgggc aaagtgtctg gggagcagct attgaccagc tacttgattt ccatccggct 1860
gttgaagaag ctcgcgaaat ggcacaacgt gtgaatcagc ccctccccgt ctggcttgct 1920
gatcgaatcc tcgaaaaaac ggcgcaacca tcagatgctg tcgccctggc gaaagggcta 1980
cacgtggtgc cggaatttct cggaatcgc gcccccttcg cagatcctca tgccagagcg 2040
gtaatttgtg gcctgggtat ggaagcagat ctggataatt tactcgctt gtatatcgct 2100
ggattatgcg gaattgggta tggctctgcg caaattctcg acgctcaaac agcgcagggg 2160
gtagttagta aaaatatcgt tattagcggc ggtgcggggc agcatccact ggtacgacaa 2220
atttggcgag atacctgctg tattccggtc attaccagc aatgctgcga accggtttta 2280
ttaggctcgg ccattcttgg tgcgtcgcc ggaaatattg caccttctgt tggcgaagcg 2340
atgcaacaat tcacccatgt ggataaatat tattatccgc aagaacgcta tcaatctctt 2400
catcatcgtc gatatgaggc ctataagcag ttgcagcata ctgcaaaatt actcagagac 2460
taattaacca gccgcctgac gctgttttca ggcaatcact aataacgact cactccggta 2520
atatcccgga gtgcattcat ctgcacccta aaaacgaggt ttatatgtcc agaaataata 2580
aacagtgggt gggtttgcca ctgcacttga tatggggata tatcgccatc gcagtattta 2640
tgactggtag tggtttcgaa ctcgcattct tatcgcacta tattaatcg ttaggcttca 2700
cacctgcgga agcctctttt gcctttacgc tctacggcct ggcggtgctc ctttccgcct 2760
gggtttctgg ggtagtagcg gaaatcatca cgccgcaaaa aacctgctg attggttttg 2820
tcctatgggt cgtattccat gttctgtttc tggctcttgg attaggacag gcaaactatg 2880
ggttaatcct gctgttttac gggattcggt gtttagcgta tccgctatct ctctactcat 2940
ttatcgctgt tattattcat aacgtgcgta gcgaaaattc cagttctgct ctgggttggt 3000
actggcggtt atattcagta gggatcggtg ttgctggcag ttatattccc agttttacga 3060
taccgattat gggtgaaatg ggaaccttat ggctggcact ggctgtctgc tttgctggcg 3120
gtgtcattgc catgatctcc ttgcgtcatg ttaaaacgcc tggacatatg cataatttaa 3180
ctccccgtga gaagtttgca gaattaagtc gggcagtaac tttactttat accaaccgca 3240
atatttttct ctccagtatt gtgcgcatta taaatacctt atcgttattc ggttttgcgg 3300
tcattatgcc aatgatgttt gtggatgaac tgggattcac cacctctgaa tggttgcagg 3360
tctgggcggc atttttcttt accactattt tctctaatat tttttggggg attgtggcag 3420
aaaaaatggg ctggatcgct gttattcgct ggtttgggtg cctcgggatg gcagcatcaa 3480
gtttagcgtt ttactacatg ccgcaataact ttggtcacaa ctactggatg gcaatgattc 3540
cggcgattgc tctgggaact tttgttgctg catttgtgcc gatggccgct gtcttcccgg 3600
cactggaacc aaaacacaaa ggtgctgcaa tctcggttta caacctctct gcgggtatgt 3660
ctaacttcct ggctccggca attgccgtgg tgttattacc gtgggttagc actatcggtg 3720
tggtcattgc ctatacagca ttgtatctat tggcctttgt cctttgcgca ttcattcgcg 3780
ttgagcagcc aggattcagt tctgcgccag tgactgagaa ggcattgaat atctcctgaa 3840
aaacgaaacg catcaggcac tcatcctctc cctcatggga gaggatgatt tcacatcagg 3900
caatagtgc tttgttatcg agataaacgt cctgcacggc gttaatcagt ttcacgcgct 3960
cagccatcga t

```

<212> Type : DNA

<211> Length : 3971

SequenceName : SEQ ID NO 2

SequenceDescription : ribitol operon

Feature

Sequence: SEQ ID NO 2:

<221> FeatureKey : misc_feature

<222> LocationFrom : 96

<222> LocationTo : 848

Other Information : ribitol dehydrogenase coding region

CDSJoin : No

Feature

Sequence: SEQ ID NO 2:

<221> FeatureKey : misc_feature

<222> LocationFrom : 859

<222> LocationTo : 2463

Other Information : ribitol kinase coding region

CDSJoin : No

Feature

```
Sequence: SEQ ID NO 2:
<221> FeatureKey : misc_feature
<222> LocationFrom : 2565
<222> LocationTo : 3839
Other Information : ribitol transporter coding region
CDSJoin : No
```

Custom Codon

Sequence Name : ribitol operon

Sequence

```
<213> OrganismName : Escherichia coli
<400> PreSequenceString :
MMNHSVPSMN TPLNGKVA AI TGAASGIGLQ CAKTL LDAGA KVVLI DREGD KLHKIVAELG      60
ENAYALQLDL FNNQQVDNML ADIIELAGGL DIFHANAGAY IGGPVAEGDP DVWDRVLNLN      120
INAAFR CVRA VLP HMIAQRS GDII FTSSIA GVVPI WEPI YTASKFAVQA FVHTTRRQVS      180
QYGV RVGAVL PGPVVTALLD DWPKAKMEEA LANGSLMQPI EVAESVLFMV TRSKNVTVRD      240
LVILPGSVDL                                     250

<212> Type : PRT
<211> Length : 250
SequenceName : SEQ ID NO 3
SequenceDescription : ribitol dehydrogenase
```

Sequence

```

<213> OrganismName : Escherichia coli
<400> PreSequenceString :
MTITKTIVIGV DVGSGSVRAG IFDLNGLLS HATEKITTR RSGSRVEQSS QEIWQAVCSC      60
IRNALTLADV CAQSVAGIGF DATCSLVVLD KNGDPLVPSP EGDAKQNIIV WMDHRATEQA      120
ERINATHHPV LNYVGGKISP EMETPKILWL KENMPEIYER AGQFFDLADF LTWRATGDLA      180
RSVCTVTCKW TWAHAENRWD PDIYFRTIGLA ELADED FIRI GHIVSPGPT CGNGLTAQAA      240
AEMGLLPGTP VAVGLIDAHA GGIGTVGVEG GALNNLAYVF GTSSCTMAST TSPSFVPGVW      300
GPYYSAMVPG LWLVEGGQSA AGAAIDQLLD FHPAVEEARE MAQRVNQPLP VWLADRILEK      360
TAQPSDAVAL AKGLHVVPEF LGNRAPFADP HARAVICGLG MERDLNLLA LYIAGLCGIG      420
YGLRQILDAQ TAQGVVSKNI VISGGAGQHP LVRQILADTC GIPVITTQCC EPVLLGSAIL      480
GAVAGNIAPS VGEAMQQFTH VDKYYYPQER YQSLHRRYE AYKQLQHTAK LLRD              534

<212> Type : PRT
<211> Length : 534
SequenceName : SEQ ID NO 4
SequenceDescription : ribitol kinase

```

Sequence

```
<213> OrganismName : Escherichia coli
<400> PreSequenceString :
MSRRNNKQWLGL LPLHLIWGYI AIAVFMGTGDG FELAFLSHYI KSLGFTPAEA SFAFTLYGLA      60
AALSAWVSGV VAEIITPQKT MLIGFVLWCV FHVLFVLVFLG GQANYGLILL FYGIRGLAYP      120
LFLYSFIVVI IHNVRSENS SALGWYWAVY SVGIGVAGSY IPSFTIPIMG EMGTLWLALA      180
FCFAGGVIAM ISLRHVKTGP MHMNLTPREK FAELSRVAVTL LYTNRNIFLS SIVRIINTLS      240
LFGFAVIMPM MFLVDELGFTT SEWLQVWAAF FFTTIFSNIW WGIVAELKMGW MRVIRWFGCL      300
GMAASSLAFY YMPQYFGHNY WMAMIPAIAL GTFVAAFVPM AAVFPALPK HKGAAISVYN      360
LSAGMSNFLA PAIAVLLPW FSTIGVVIAY TALYLLAFVL CAFIRVEQPG FSSAPVTEKA      420
LNIS                                                                    424

<212> Type : PRT
<211> Length : 424
SequenceName : SEQ ID NO 5
SequenceDescription : ribitol transporter
```